Activity-aware map: identifying human daily activity pattern using mobile phone data

Conference or Workshop Item

Activity-Aware Map: Identifying Human Daily Activity Pattern Using Mobile Phone Data

Santi Phithakkitnukoon[1], Teerayut Horanont[1,2], Giusy Di Lorenzo[1], Ryosuke Shibasaki[2], and Carlo Ratti[1]

[1] SENSE*able* City Laboratory, School of Architecture and Planning, Massachusetts Institute of Technology, Cambridge, MA, USA

[2] Department of Civil Engineering, School of Engineering, The University of Tokyo, Tokyo, Japan

{santi,giusy,ratti}@mit.edu, teerayut@iis.u-tokyo.ac.jp, shiba@csis.u-tokyo.ac.jp

Abstract. Being able to understand dynamics of human mobility is essential for urban planning and transportation management. Besides geographic space, in this paper, we characterize mobility in a profile-based space (*activity-aware map*) that describes most probable activity associated with a specific area of space. This, in turn, allows us to capture the individual daily activity pattern and analyze the correlations among different people's work area's profile. Based on a large mobile phone data of nearly one million records of the users in the central Metro-Boston area, we find a strong correlation in daily activity patterns within the group of people who share a common work area's profile. In addition, within the group itself, the similarity in activity patterns decreases as their work places become apart.

1 Introduction

For better understanding of the effects of human movement, characterizing human mobility patterns is crucial. For example, without such characterization, the impact of inhabit dynamics in the city cannot be understood. As spatio-temporal and geo-referenced datasets are growing rapidly because of the daily collection of transaction data through database systems, network traffic controllers, sensor networks, and telecommunication data from mobile phones and other location-aware devices, the large availability of these forms of data allows researchers to better characterize human mobility. The additional information of activities associated with human mobility further provides a unique opportunity to better understand the context of human movement, and hence better urban planning and management. In this paper, we develop the *activity-aware map*, which provides information about the most probable activity associated with a specific area in the map. With the activity-aware map and an analysis of a large mobile phone data of nearly one million records of location traces, we are able to construct the individual daily activity patterns. This allows us to carry out a correlation analysis of work area's profile and similarity in daily activity patterns.

A.A. Salah et al. (Eds.): HBU 2010, LNCS 6219, pp. 14–25, 2010.

2 Related Work

A rapidly increasing number of mobile phone users has motivated researchers from various fields to study its social [1][2][3] and economic [4][5][6] impact. With the extensive records of mobile phone data such as calling pattern and location of the mobile phone user, analyses have been performed on numerous aspects including behavioral routine [7][8][9], social proximity [10][11], call prediction [12][13], social closeness [14][15], and human mobility [16][17][18][19][20].

Understanding dynamics of social networks is beneficial to urban planning, public transport design, traffic engineering, disease outbreaks control, and emergency response management. To study dynamics in human mobility, GPS receiver has been handy for researchers in collecting a large real-life traces. Azevedo et al. [16] study pedestrian mobility behavior using GPS traces captured at Quinta da Boa Vista's Park in Rio de Janeiro (Brazil). Movement elements are analyzed from data collected from 120 pedestrians. They find that the velocity and acceleration elements follow a normal distribution while the direction angle change and the pause time measure fit better to lognormal distribution. Based on 226 daily GPS traces of 101 subjects, Lee et al. [17] develop a mobility model that captures the effect of human mobility patterns characterized by some fundamental statistical functions. With analytical and empirical evidence, they show that human movement can be expressed using gaps among *fractal waypoints* [21] (people are more attracted to more popular places).

With a large set of mobile phone data, Candia et al. [18] study spatiotemporal human dynamics as well as social interactions. They investigate the patterns in anomalous events, which can be useful in real-time detection of emergency situation. At the individual level, they find that the interevent time of consecutive calls can be described by *heavy-tailed* distribution, which is consistent with the previous reports on other human related activities. Gonzalez et al. [19] examine six-month trajectory of 100,000 mobile phone users and find a high regularity degree in human trajectories contrasting with estimation by *Levy flight* and *random walk* models. People tend to return a few frequent locations and follow simple repeated patterns despite the diversity of the their travel history. The most recent study in human mobility based on a large mobile phone data by Song et al. [20], whose result is consistent with Gonzalez et al.'s [19] that human mobility is highly predictable. Based on data from 50,000 mobile phone users, they find that predictability in human mobility is independent of distance that each individual regularly travel and show that the predictability is stabled at 93% for all regular traveled distances of more than 10km.

In contrast with other work in human mobility, our work is focusing on human mobility concerning the spatial profile (i.e. type of space or surrounding area such as dinning, shopping, and entertainment) rather than geographical location.

3 Methodology

A number of literature have described geographical human mobility pattern concerning movement of people between multiple locations. Here we are interested in characterizing the mobility not by geographic location but its associated *spatial profile*. This spatial profile-based mobility pattern, in turn, becomes a *human activity pattern*. In addition, our interest expands to investigation of relationship between this activity pattern and demographic of people. Therefore, in this section, we will describe our methodology used in characterizing space, capturing daily activity pattern, as well as preprocessing our dataset.

3.1 Data Preparation

In this research, we use anonymous mobile phone data collected during the period from July 30th, 2009 to September 12th, 2009 by Airsage[22] of about one million users in the state of Massachusetts, which account for approximately 20% of population, equally spread over space. This includes 130 million anonymous location estimations in (latitude,longitude)-coordinates, which are recorded when the users are engaged in communication via the cellular network. Specifically, the locations are estimated at the beginning and the end of each voice call placed or received, when a short message is sent or received, and while internet is connected. Note that these location estimations have an average uncertainty of 320 meters and median of 220 meters as reported by Airsage[22] based on internal and independent tests. For our analysis, we consider the mobile phone data within an area of 33x42km^2, which includes 52 cities (Boston, Cambridge, and others) in the county of Essex, Middlesex, Suffolk, and Norfolk as shown in Fig. 1. The list of the counties and their corresponding area covered (in km^2) by this study are shown in Table 1.

Within this area in the map, we need to extract mobility traces of each user from the mobile phone data. As the estimation of the user's location is aggregated

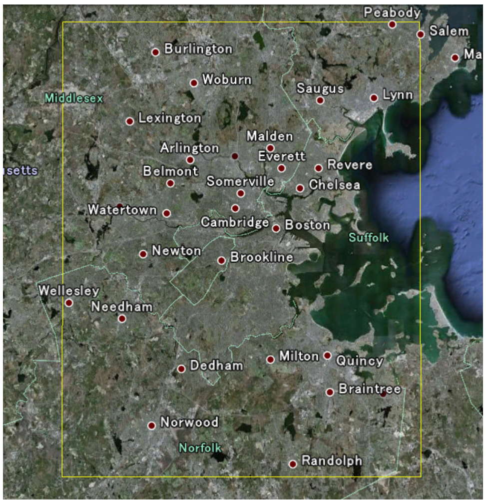

Fig. 1. Area of study in this research, cropped by yellow line

Table 1. List of the counties and their area covered by this study

County	Area covered (km^2)
Essex	110.30
Middlesex	452.52
Suffolk	154.39
Norfolk	26.12

only when network connection is established, mobility thus can be derived as a temporal sequence of locations. To segment these traces into trajectories so that daily mobility pattern of each individual can be identified, we describe here some basic algorithms to extract *trajectory* and *stop* [23].

Let X_k denote a set of sequential traces of user k such that $X_k = \{x_k(1), x_k(2), x_k(3), ...\}$ where $x_k(i)$ is a position i of user k. A trajectory can then be obtained by segmenting X_k with the spatial threshold $\triangle S$. If a distance between adjacent positions is greater than the threshold ($distance(x_k(i), x_k(i+1)) > \triangle S$), then the early position $x_k(i)$ becomes the end position of the last trajectory while the later position $x_k(i+1)$ becomes the starting position of the next trajectory. Once the trajectories are detected, a stop can be identified as an event during which the user stays in a specific location for a sufficiently long period of time. As each position i contains location and timestamp, i.e. $x_k(i) = (lat(i), long(i), t(i))$, extraction of a stop depends on time and space. A stop is thus regarded as a sequence of positions $\{x(j), x(j+1), x(j+3), ..., x(j+m)\}$ where the distance between any adjacent positions is less than a spatial threshold S_{th} i.e., $distance(x(j), x(j+1)) < S_{th}$, and time spent within the location is greater than a time threshold T_{th} i.e., $t(m) - t(j) > T_{th}$.

After stops have been identified, *work* location of each user is then estimated as a most frequent stop during the day hours. The information about work location allows us to derive the mobility choices of the users, and detect activity patterns throughout the day.

3.2 Spatial Profiling

To model the space, we construct a virtual grid reference by dividing the map into square cells of size 500 by 500 meters (to compensate location estimation uncertainty). Since our interest is in the activities associated with the space, we thus characterize space based on the type of activities expected to be performed within given space. For example, if restaurants were clustered within a particular area, then this area would be associated with eating activity.

In this study, we consider four different human activities in which people typically spend time engaging on daily basis. These activities are concerning eating, shopping, entertainment, and recreational. Profiling the map according to these activities requires information about the types of places within each cell. To acquire the information regarding these activities, we search for Points of Interest

Table 2. Considered activities and keywords used for POIs search

Activity	Keywords used
Eating	Restaurant, Bakery, Coffee shop
Shopping	Mall, Store, Market
Entertainment	Theater, Bowling, Night club
Recreational	Park, Gym, Fitness

(POIs) for each cell location. We use *pYsearch* (Python APIs for Y! search services) version 3.1 [24] for POI search service, and Reverse Geocoding with *Geopy* (A Geocoding Toolbox for Python) [25] for translating $(latitude, longitude)$-coordinate into a physical address. For each activity category of each cell, we make three search attempts using different keywords. The keywords used for each activity category are listed in Table 2. With the limit of 5,000 queries per day restricted by Yahoo, an extensive amount of search time is required inevitably.

Once POI searches are completed, the number of POIs associated with each activity category is recorded for each cell. The raw *activity distribution map* is then composed of 500x500m^2 cells where each cell contains distribution of each activity. Each cell C_i contains normalized portion of each activity:

$$C_i = [\alpha_i(1), \alpha_i(2), \alpha_i(3), \alpha_i(4)], \tag{1}$$

where $i = 1, 2, 3, ..., N$, N is the total number of cells, and normalized portion of each activity $\alpha_i(a)$ in cell i is computed as

$$\alpha_i(a) = \frac{n_{\alpha_i(a)}}{\sum_{i=1}^{N} n_{\alpha_i(a)}}, \tag{2}$$

where $n_{\alpha_i(a)}$ denotes the number of POIs associated with activity a within the cell i and $a = 1, 2, 3, 4$ corresponds to eating, shopping, entertainment, and recreational activity, respectively.

Based on our POI search, Fig. 2 shows a map with the visual grids and POIs found by 12 different keywords (described in Table 2) in different colors.

To further classify these cells into a more crisp distribution map, we apply k-means algorithm with k=4. The resulting *crisp activity distribution map* is depicted in Fig. 3 where each cell is classified to one of the four activities according to Bayes theorem:

$$P(a|n_{\alpha_i(a)}) = \frac{P(n_{\alpha_i(a)}|a)P(a)}{n_{\alpha_i(a)}}. \tag{3}$$

The interest here is to find the most probable activity category a for each of the k clusters. Therefore, for each cluster, we find a that maximizes *a posteriori* (MAP method). So we use Bayes theorem above to compute the posterior probability of each activity category as follows:

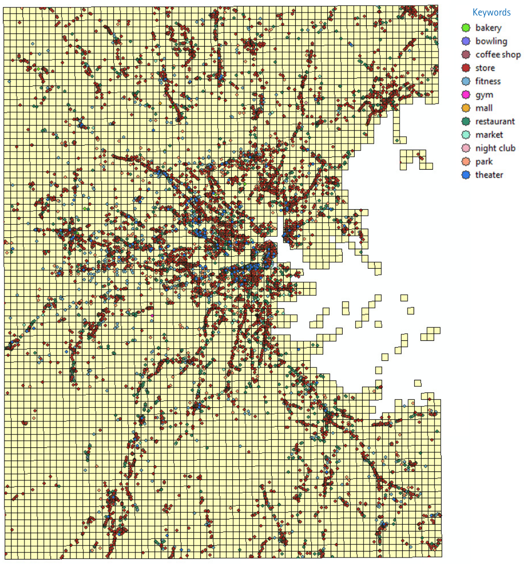

Fig. 2. POI search results on the map with 500x500m^2 visual grids

$$\begin{aligned} a_{MAP} &\equiv \arg\max_a P(a|n_{\alpha_i(a)}) \\ &= \arg\max_a \frac{P(n_{\alpha_i(a)}|a)P(a)}{n_{\alpha_i(a)}} \\ &= \arg\max_a P(n_{\alpha_i(a)}|a)P(a). \end{aligned} \tag{4}$$

3.3 Daily Activity Patterns

Generally, people perform different activities throughout the day. A lot of these activities are repeated on daily basis, e.g. eating around 12pm (noon), jogging in the evening, and hence producing recognizable patterns. With our mobile phone data, each user is more likely to engage in an activity during "stop" rather than on the move. Therefore, for each stop, activity is identified according to the crisp activity distribution map.

To infer a daily activity pattern for each user, we divide 24-hour time scale into eight 3-hour segments starting at 5AM as shown in Fig. 4. So daily activity pattern is simply a sequence of activities performed by the user during each stop throughout the day. For each user, daily activity patterns are collected over the course of the data collection period. Note that, in this study, we consider only weekdays (Monday, Tuesday, Wednesday, Thursday, Friday) as our speculation is that weekday pattern is different from weekend pattern due to typical work schedule and hence different daily activity sequences – this will be addressed and further discussed in our future work.

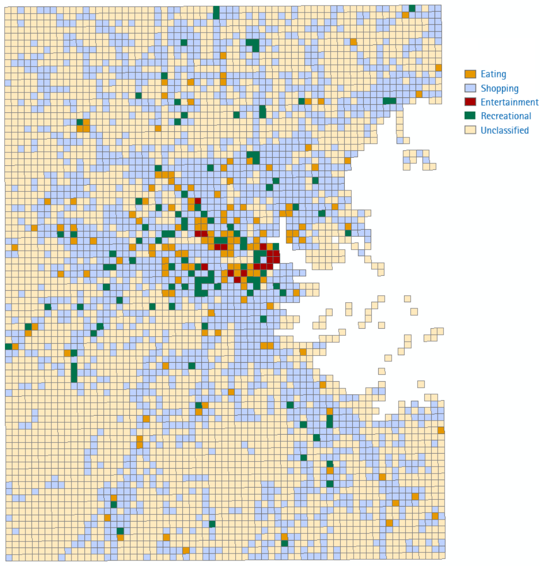

Fig. 3. Crisp activity distribution map

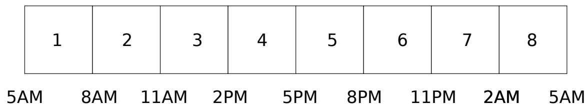

Fig. 4. The eight 3-hour temporal windows are used to frame the daily activity pattern

To derive the representative daily activity pattern of each user, we simply assign each segment with the most frequent activity during that time interval over the period of data collection. Precisely, if $\lambda_a^d(t)$ represents the count of activity a on d-th day during time segment t (where $t = 1, 2, 3.., 8$), then

$$z(t) = \arg\max_a \sum_{d=1}^{M} \lambda_a^d(t) \tag{5}$$

where $z(t)$ is the assigned activity for time segment t and M is the total number of days.

4 Work Area's Profile and Similarity in Daily Activity Patterns

The activity map and individual daily activity patterns developed in the previous section allows us to conduct a number of studies that can be useful for better understanding of human behavior in the city. In this present research, we are particularly interested in relationship between people's daily activity patterns and the characteristic of their work area. Do people who work in the same area's category (e.g. eating, shopping, etc.) also have similar daily activity patterns? With the same type of work area, how does distance impact the similarity in their daily activity patterns (e.g. do people who work in an urban shopping area have similar activity pattern with people who work in a distant shopping area)? In this current study, we are attempting to answer these two questions.

As a first step, we classify the users into four groups based on their work cell's profiles. Each group then consists of a number of different individual daily activity patterns who have a common work cell's profile. To represent each group's activity pattern, we need to find a group signature for further correlation analyses. The representative daily activity pattern or signature of each group can be obtained in a similar fashion with the individual patterns described in the previous section (using Eq. (5)). The derived signatures are shown in Table 3.

It can be noticed that there is no Eating element appears in any of other group signatures beside its own group (showing in form of a working activity, W). Our speculation is that it could be caused by first, people normally eat at home (breakfasts) and at work or somewhere nearby workplace (lunches), and second, people are not frequently involved in a phone communication while at eating area. Note also that the patterns are derived from weekdays activities so if weekends-only activities are considered, Eating elements could emerge in the group patterns.

To answer the first question, we need to measure similarity in daily activity patterns among individuals within the same group as well as among other groups. To measure distance (dissimilarity) between two daily activity patterns, we use *Hamming distance*, which is normally used to measure distance between two strings of equal length. The distance is essentially the number of positions at which the corresponding symbols are different, which is quite suitable for our case as a series of activities can be considered as symbols. The result of the average Hamming distance within the group is shown in Table 4.

Using group signatures obtained earlier, we then measure dissimilarity between each group signature and other group's individual patterns. The result of this between-group distance is shown in Table 5 in forms of average Hamming distance.

As the result of our first investigation, Fig. 5 illustrates a bar plot intended to make a comparison between within-group and between-group distances where

Table 3. Signature of each group based on work cell's profile. Note: Eat. = Eating, Sho. = Shopping, Ent. = Entertainment, Rec. = Recreational, W = Work cell.

Group	Group's daily activity pattern
Eating	W–W–W–W–Sho.–Rec.–Rec.–Sho.
Shopping	W–W–W–W–Rec.–Rec.–W–W
Entertainment	Sho.–W–W–W–W–Rec.–Sho.–Sho.
Recreational	W–W–W–W–W–Sho.–Sho.–Sho.

Table 4. Average within-group distance

Work cell's profile	Average distance
Eating	4.78
Shopping	2.58
Entertainment	4.67
Recreational	3.61

Table 5. Average between-group distance

	Eating	Shopping	Entertainment	Recreational
Eating	–	6.53	6.60	6.96
Shopping	4.90	–	4.92	5.05
Entertainment	6.43	6.88	–	7.00
Recreational	5.04	4.81	5.13	–

red bars represent within-group distance while blue bars represent between-group distance. Clearly, it shows that within-group distances are less than between-group distances. This implies that people who have a common work cell's profile tend to exhibit more similar daily activity patterns than people who have different work cell's profile.

For the second investigation about the impact of physical distance on the similarity in activity patterns, we decide to proceed by placing a growing spatial window (a circle of an arbitrary radius) onto the map then measure similarity between between the users' activity pattern whose work cell located at the center of the window and other users whose work cells are within the vicinity of the spatial window. The similarity is being measured while the radius of the window grows from a small to larger value. The process is repeated for each activity category. This way, we can see the change(if any) in similarity for each work profile as we move away from the center area. Precisely, we choose to grow the spatial window from the center of Boston area with the radius varying

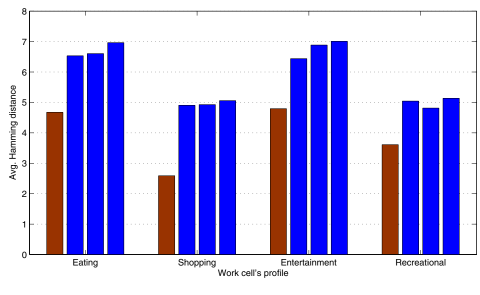

Fig. 5. When users are grouped together based on their work cell's profiles, within-group and between-group distances are illustrated with red and blue bars respectively. This shows higher degree in similarity within the group than between groups.

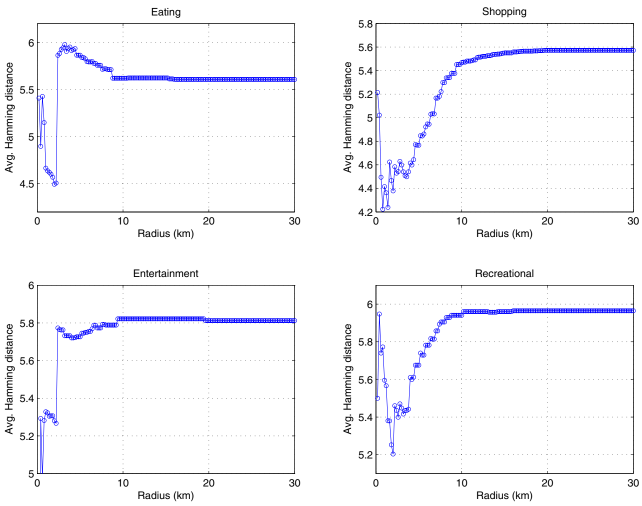

Fig. 6. Dissimilarity in daily activity patterns is measured by average Hamming distance as the radius varies from 0.5km to 30km for each work cell's profile. The center of the growing radius is near the center of city of Boston. Dissimilarity is between the users whose work cells are within the 0.5km radius and other users covered by growing radius.

from 0.5km to 30km. The result for each work category is shown in Fig. 6. We can observe that, overall, the similarity in activity patterns decreases as radius increases, which implies that physical distance has an impact on similarity in daily activity patterns. People whose work area's profile are although the same, their activity patterns tend to deviate more as they work areas become further apart.

In summary, we have observed a strong correlation in daily activity patterns within the group of people who share a common work area's profile. Addition, within the group itself, the similarity in activity patterns decreases as the distance between them increases.

5 Limitations of the Study

There are a number of limitations of this study. First and foremost, the lack of continuity of mobility traces due to the fact that the location is estimated from mobile phone data only when connection with a cellular network is made through either voice, text, or data communication, which constricts us to a smaller number of users that can be analyzed. Secondly, our POI search is constrained by Yahoo's search limit and capability. Lastly, home and work locations are estimated intuitively according to the data provided. Although ground-truth validation is desired, it would be very difficult to perform due to the privacy issue.

6 Conclusions

In this paper, we have developed an activity-aware map that contains most probable activity associated with a specific area in the map based on POIs information. With activity-aware map, we are able to extract individual daily activity patterns from analyzing a large mobile phone data of nearly one million records. Results from our correlation analysis show a strong correlation in daily activity patterns within the group of people who share a common work area's profile. In addition, within the group itself, the similarity in activity patterns decreases as the distance between them increases. This study is the first report of many more to come in using activity-aware map to study inhabitant behavior. So as our future direction, we will continue to investigate on daily activity pattern and its dynamics for better understanding of human dynamics, which in turn benefits urban planning and management.

Acknowledgments. This research has been supported by VWoA Electronic Research Lab, AT&T Foundation, National Science Foundation, MIT Portugal, and all the Senseable City Laboratory Consortium members. We would like to also express our gratitude to Aisage for valuable data.

References

1. Turner, M., Love, S., Howell, M.: Understanding emotions experienced when using a mobile phone in public: The social usability of mobile (cellular) telephones. Telemat. Inf. 25(3), 201–215 (2008)
2. Nickerson, R.C., Isaac, H., Mak, B.: A multi-national study of attitudes about mobile phone use in social settings. Int. J. Mob. Commun. 6(5), 541–563 (2008)
3. Liu, C.C.: Measuring and prioritising value of mobile phone usage. Int. J. Mob. Commun. 8(1), 41–52 (2010)
4. Kauffman, R.J., Techatassanasoontorn, A.A.: International diffusion of digital mobile technology: A coupled-hazard state-based approach. Inf. Technol. and Management 6(2-3), 253–292 (2005)
5. Giray, F., Gercek, A., Oguzlar, A., Tuzunturk, S.: The effects of taxation on mobile phones: a panel data approach. Int. J. Mob. Commun. 7(5), 594–613 (2009)
6. Li, W., McQueen, R.J.: Barriers to mobile commerce adoption: an analysis framework for a country level perspective. Int. J. Mob. Commun. 6(2), 231–257 (2008)
7. Eagle, N., Pentland, A.: Reality mining: sensing complex social systems. Personal and Ubiquitous Computing 10(4), 255–268 (2006)
8. Eagle, N., Pentland, A.: Eigenbehaviors: Identifying structure in routine. Proc. Roy. Soc. A (2006) (in submission)
9. Eagle, N.: Machine perception and learning of complex social systems. Ph.D. Thesis, Program in Media Arts and Sciences, Massachusetts Institute of Technology (2005)
10. Clauset, A., Eagle, N.: Ersistence and periodicity in a dynamic proximity network. In: Proceedings of Discrete Mathematics and Theoretical Computer Science Workshop on Computational Methods for Dynamic Interaction Networks (2007)
11. Eagle, N., Pentland, A., Lazer, D.: Inferring social network structure using mobile phone data. PNAS (2007)

12. Phithakkitnukoon, S., Dantu, R.: Predicting calls — new service for an intelligent phone. In: Krishnaswamy, D., Pfeifer, T., Raz, D. (eds.) MMNS 2007. LNCS, vol. 4787, pp. 26–37. Springer, Heidelberg (2007)
13. Phithakkitnukoon, S., Dantu, R.: Cpl: Enhancing mobile phone functionality by call predicted list. In: Meersman, R., Tari, Z. (eds.) OTM 2008, Part II. LNCS, vol. 5332, pp. 571–581. Springer, Heidelberg (2008)
14. Phithakkitnukoon, S., Dantu, R.: Mobile social group sizes and scaling ratio. AI & Society, Springer (2009)
15. Phithakkitnukoon, S., Dantu, R.: Mobile social closeness and similarity in calling patterns. In: IEEE Conference on Consumer Communications & Networking Conference (CCNC 2010), Special Session on Social Networking, SocNets (2010)
16. Azevedo, T.S., Bezerra, R.L., Campos, C.A.V., de Moraes, L.F.M.: An analysis of human mobility using real traces. In: WCNC 2009: Proceedings of the 2009 IEEE conference on Wireless Communications & Networking Conference, Piscataway, NJ, USA, pp. 2390–2395. IEEE Press, Los Alamitos (2009)
17. Lee, K., Hong, S., Kim, S.J., Rhee, I., Chong, S.: Slaw: A mobility model for human walks. In: Proceedings of the 28th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), Rio de Janeiro, Brazil. IEEE, Los Alamitos (April 2009)
18. Candia, J., Gonzalez, M.C., Wang, P., Schoenharl, T., Madey, G., Barabasi, A.: Uncovering individual and collective human dynamics from mobile phone records. Journal of Physics A: Mathematical and Theoretical 41(22), 1–16 (2008)
19. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L.: Understanding individual human mobility patterns. Nature 453(7196), 779–782 (2008)
20. Song, C., Qu, Z., Blumm, N., Barabasi, A.L.: Limits of predictability in human mobility. Science 327(5968), 1018–1021 (2010)
21. Rhee, I., Lee, K., Hong, S., Kim, S.J., Chong, S.: Demystifying the levy-walk nature of human walks. Technical report, NCSU (2008)
22. Airsage: Airsage wise technology, `http://www.airsage.com`
23. Calabrese, F., Pereira, F.C., Lorenzo, G.D., Liu, L.: The geography of taste: analyzing cell-phone mobility and social events. In: Proceedings of IEEE Inter. Conf. on Pervasive Computing, PerComp. (2010)
24. pYsearch: Python APIs for Y! search services, `http://pysearch.sourceforge.net/`
25. Geopy: A Geocoding Toolbox for Python, `http://code.google.com/p/geopy/wiki/ReverseGeocoding`